

# 评价连续体:大概念教学的评价设计\*

徐玲玲 刘 徽 曹 琦

〔摘 要〕大概念教学要真正落实,必须由评价予以保证。大概念教学的评价设计最终指向素养,而素养目标的高阶性和目标构成的复杂多元性,决定了大概念教学的评价应当是一个连续体。评价连续体由多元的评价类型和评价方法构成,评价类型包括学习性评价、学习的评价和学习式评价,评价方法包括课堂问答、知识和技能测验、表现性任务、结构化思维工具、自我反思等,最终的评价任务多表现为真实性问题情境,同时反映素养和大概念,从而保证素养目标的达成。评价连续体表现为评价类型的连续体和评价方法的连续体。

〔关键词〕大概念;评价;连续体

DOI:10.16194/j.cnki.31-1059/g4.2022.01.004

大概念教学的目标指向素养,素养目标的构成是复杂多元的,这就要求其评价设计遵循整合性的评价逻辑。所谓整合性评价逻辑,与终结性评价逻辑不同,它引入了复杂情境,强调在真实性问题情境中考量学生的素养表现,“只有真实情境下运用某种或多种知识完成特定的任务,才能评估关键能力、必备品格与价值观念”。<sup>[1]</sup>

在整合性评价逻辑下,表现性评价(performance assessment)是最为常见的评价方式之一,表现性评价中的表现性任务(performance task)也即真实性问题情境三要素中的“任务”。好的评价体系的标准是“恰当的,有效的,可信的,行之有效的,可以实现的”,<sup>[2]</sup>这也决定了评价类型和评价方法的多元,具体呈现为琳达·达令-哈蒙德和弗兰克·亚当森所说的“评价连续体(assessment continuum,也译为评价连续统)”。<sup>[3]</sup>

## 一、确定最终的评价任务

(一)评价的三种类型:学习性评价、学习的评价和学习式评价

过去,我们在给评价分类时,应用最多的是形成性评价和终结性评价。然而,对形成性评价和终结性评价普遍存在的一个误解是,认为形成性评价是在学习过程中进行的评价,终结性评价则是在学习结束后进行的评价。实际上,形成性评价和终结性评价的本质区别并非其实施的阶段不同,而在于不同的评价目的——形成性评价旨在提供反馈,终结性评价则是为了评定水平。除上述分类之外,Earl依据评价与学习的关系将评价分为三种类型:一是学习性评价(assessment for learning),指为了推进学习而进行的评价;二是学习的评价(assessment of learning),指为了评定学习水平而进行的评价;三是学习式评价(assessment as learning),指为了让学生在学中评价而进行的评价。<sup>[4]</sup>

\* 本文系2016年国家社会科学基金教育学一般项目《中小学课堂学习环境的设计研究》(编号:BHA160093)的阶段性研究成果。通讯作者:刘徽。

这三类评价和形成性评价、终结性评价有一定关联,也存在区别。从评价目的看,“学习性评价”和“学习的评价”的关系类似于“形成性评价”和“终结性评价”的关系。而“学习式评价”则是原有的形成性评价和终结性评价这种分类方式所忽略的。事实上,在真实世界中,每个人都需要学会评价,不仅仅要会评价他人,更重要的是会正确地评价自我。与形成性评价、终结性评价相比,Earl的这种评价分类法更清晰地呈现了评价目的。后来,Stern等将这三种类型应用于概念为本的教学中。<sup>[5]</sup>

评价目的的不同决定了评价标准的侧重点不同。“学习性评价”的标准侧重于“具体”,这一评价类型的目的在于为学习的推进提供依据,所设计的学习性评价越具体,越能促进学习。教师在学习前、学习中和学习后及时收集学习证据,同时提供反馈,帮助学生更好地改进学习。“学习的评价”的标准侧重于“公平”,这一评价类型旨在根据阶段性的学习成果对学生的学习水平进行评定、分类和筛选。教师根据所收集到的信息和证据总结学生在特定阶段的学习情况,公平公正地作出评定,并向学生、家长、教师和社会等传达结果。“学习式评价”的标准侧重“自省”,这一评价类型重点关注学生评价能力的培养,指在教师的支持、示范及指导下,学生学会收集相应的证据,不仅为教师和同伴提供评价的信息,更重要的是在这个过程中进行自我评价,从而反思学习过程、调整学习方法,同时设立更为精准合理的学习目标。<sup>[6]</sup>

表1 三种类型的评价

评价类型	内涵	样例 <sup>①</sup>	评价标准
学习性评价	为了推进学习而进行的评价	请寻找现实生活中的三种统计图,并运用统计图的知识加以解释	具体
学习的评价	为了评定学习水平而进行的评价	情境任务:你作为一名市场运营专员,需要在公司会议上汇报2021年度的市场运营情况,并对2022年的市场运营做一个预估分析。要求学生: (1)绘制合适的统计图表,并配上相应文字用以呈现年度市场运营情况; (2)撰写预估分析报告,根据上一年度运营情况做出下一年度的分析预测	公平
学习式评价	为了让学生在中学学会评价而进行的评价	全班讨论设计学校生源数据的统计图表,并形成相应的文字报告,拟放在学校生源统计报告中,每位同学作为评委为他人和自己的图表报告进行评价,并结合意见完善自己的成果	自省

在实际操作时,教师很容易混淆上述三种评价的标准。例如,在制作一分钟计时器时,科学教

师提供了四种材料让每组学生选择,并根据所选材料制作不同材质的计时工具,分别是火钟、水钟、沙钟、机械钟,同时,告知学生制作完成后要对成果进行检验,将制作出的计时工具更为精准的小组评选为最优小组。这一评价类型属于“学习的评价”。在制作过程中,每个小组都想获得最优,于是学生都想选择机械钟,这样一来机械钟的制作材料就不够了,几个小组改选了其他材料。不难料想到,最后制作机械钟的小组胜出。而“学习的评价”的评价标准是“公平”,但上述四种材料对制作计时工具而言本身就存在优劣,且材料并非学生完全自愿选择。针对这一评价案例,可以有两种改进方式,一种是将之转为学习性评价或学习式评价,不进行评定,而是通过教师点评和同伴点评收集反馈,促进反思;另一种是提供充足的材料供学生自选,从而促进评价结果更为公平。

### (二)设计有效的最终评价任务

事实上,在真实教学实践中,许多评价任务考量的仅仅是“识记”而非“理解”,因此,可能存在“学生并不理解但依然可以答对题”的情况。基于此,威金斯和麦克泰格曾提出教师需要思考的一个评价之问,即“当学生并未真正掌握或理解问题涉及的内容时,是否有可能在评价方面仍然表现良好”。<sup>[7]</sup>例如,学生能够画出“ $y=2x$ ”的函数图,是否一定代表他们已经完全理解了线性关系?答案是否定的。因为学生完全可能以“照葫芦画瓢”的方式画出函数图,而对其内涵一无所知。因此,更好的评价是为学生提供一组或几组实际数据,让他们去发现其中是否具备线性的函数关系,并作图。设计单元的最终评价任务非常重要,最终评价任务的设计会影响目标的设计,换言之,目标设计和评价设计是互相修正的。

#### 1. 最终评价任务应该包含真实性问题情境

在整合性逻辑下,评价学生是否真正理解要看他们能否将大概念应用于解决真实性问题。因此,威金斯和麦克泰格主张将评价设计前置,<sup>[8]</sup>即让评价设计与目标设计一致。一般来说,最终的评价任务是包含真实性问题情境的表现性任务。真

实性问题情境可以防止低通路迁移。例如,若语文作文题目只是宽泛地要求学生“记一件难忘的事”等,就容易出现套题的情况,学生会倾向于背诵一些范文或模版来对付考试。而若设计一个真实性问题情境,例如,提供一张照片,让学生为一个民宿撰写一篇宣传文案,这就要求学生在对具体题目进行分析的基础上进行构思和写作。

## 2. 表现性任务要精准指向素养目标和大概念

在评价设计中,表现性任务对应的是本单元的素养目标,同时体现对大概念的理解。教师们在设计表现性任务时常常会出现偏差,表现为内容设计的偏差和形式设计的偏差。以统编版八年级上册语文“传记”单元为例,该单元的大概念是“传记往往记录那些在历史进程上产生重大影响的人物生平,反映人与时代的复杂互动,在真实性的基础上生动展现人物的一生”,素养目标是“能理解传记文体的价值和意义,并学会鉴赏传记”。原先教师设计的表现性任务如下:

《美丽的颜色》居里夫人的历史性时刻是发现镭,《伟大的悲剧》斯科特一行的历史性时刻是即使弹尽粮绝依然记录下最后的荣耀。同学们,你们的历史性时刻是什么?请撰写一篇自己的自传,同学们的自传将择选收录到学校校刊的《星耀人生》人物传记合集中,请结合所读、所学、所知,完成任务。

上述任务看起来很贴近学生的生活,但实际上“写自传”这一评价方式并不能充分体现学生对传记的理解。传记不同于普通的叙事,让一个初中生写自传,容易变成一篇记叙文写作。在大概念视角下,该单元的表现性任务可以修改为:

2022年,第19届亚运会将在杭州举办。随着亚运会的脚步慢慢临近,亚运氛围越来越浓厚,城市的街头巷尾也出现了许多亚运元素,如会徽、口号、吉祥物、运动图标等,古韵杭州焕发出了独特的新魅力。备战亚运会的运动员们也在自己的领域内刻苦训练、蓄势待发,希望在亚运盛会上为国争光。为了让大家更多地了解亚运会运动员,杭州亚运会组委会决定出版《星耀亚运》人物传记集。你作为《星耀亚运》(浙江版)的采编小记者,请选择一名浙江的亚运会冠军,查阅搜集相关资料(基本简介、运动经历、参加亚运会情况、赛场之外的生活等),若能采访本人则更好,为其撰写一篇传记。<sup>②</sup>

除了设计的内容要对应大概念外,形式的设计也很重要。以统编版三年级下册语文“整合信息”单元为例,该单元的大概念是“整合信息是将多种形态的信息进行分类、比较、补充和衔接,并在此基础上梳理出线索”,其素养目标为“初步学习整合信息,介绍一种事物”。原先教师设计的表现性任务为“请你搜索网络材料,介绍国宝大熊猫”。这一任务看起来似乎没有问题,但有些学生会倾向于选择一篇比较权威和全面的网络材料作为参考,这样做的话,他们就不需要经历“整合信息”的过程,就偏离了单元目标。实际上,这一表现性任务指向于“能有效评价信息,并对材料进行缩写”这一素养目标,聚焦的大概念也不同。因此该单元的表现性任务主要需要修改形式设计,换为教师为学生提供多样化资料,留给学生整合的空间,可以修改为:

2022年的冬奥会将在北京举办。憨态可掬的吉祥物“冰墩墩”一亮相,就受到了各国人民的喜爱。该吉祥物以大熊猫为原型进行设计创作,将大熊猫的形象与富有超能量的冰晶外壳相结合,体现了冬季冰雪运动和现代科技的特点。你作为一名冬奥会的志愿者,需要制作一张大熊猫的文化名片,向国际上的小朋友详细介绍国宝大熊猫,让他们知道为何我们喜欢把大熊猫作为吉祥物。

张老师给大家准备了一些资料,你可以从国宝地位、可爱外形、日常生活、美好寓意、过去的大熊猫吉祥物形象等方面来整合信息,当然也可以增加自己的补充,让这个可爱的动物形象作为文化符号,讲述中国故事。<sup>③</sup>

## 二、设计评价连续体

尽管在整合性逻辑下,大概念教学的评价最终指向于考量学生能否在一个真实性情境中解决问题,但这并不是说大概念教学的评价只有表现性评价这一种形式,而是需要根据不同的评价目标选择不同的评价类型和方法,即大概念教学的评价是一个连续体,即“评价连续体”。如琳达·达令-哈蒙德所言:“评价策略可以看作是伴随着一个连续体而存在。一端是传统考试中的选择题和封闭式题目。这些题目测试记忆和识别,但不能测量高阶思维技能或者应用能力。另一端是要求大



量的学生设计、观点和表现的评价,尤其是发掘大学和职业所需的规划和管理技能。”<sup>[9]</sup>威金斯和麦克泰格也强调,评价要从“快照”隐喻转为“剪贴簿”隐喻,“有效的评估不是一张快照,更像是收集了纪念品和图片的剪贴簿”。<sup>[10]</sup>

### (一)评价的不同方法

大概念教学的素养目标由素养、大概念、情感维目标、认知维目标(高阶认知、低阶认知、元认知)和技能维目标(复合技能、单一技能)等构成。这些目标有时会被整合评价,但有时也可以分开评价。

评价的方法是多种多样的,包括课堂问答、知识和技能测验、表现性任务、结构化思维工具(如KWL调查、思维导图、个人意涵图、概念地图)、自我反思等。这些评价的方法和评价的目标有一定的对应关系(见图1),当然这种对应关系也不是绝对的。



图1 评价目标和评价方法

#### 1.课堂问答

课堂问答是最常见的一种评价方法,因为问答是课堂中占比很大的一种活动,通过师生的口头交流,教师能及时全面地了解学生的理解和掌握程度。课堂问答对应的目标跨度很大,既可以是素养、大概念(如“为什么会有传记这种文体”“计时工具是越精准越好吗”),也可以是单一技能和低阶思维(如“‘deal’的过去时怎么写”)。课堂问答一般用于学习性评价。

#### 2.知识测验和技能测试

知识测验一般是针对低阶思维的评价方法,其形式多样,但以封闭型试题(如填空题、选择题、连线题等)为主,一般用于评价学生的知识掌握情况,如将不同的说明方法与对应的例子连线。而技能测试一般是针对单一技能的评价方法,知识测验和技能测试一般用于学习的评价。

#### 3.表现性任务

表现性任务是大概概念教学的评价设计里最常出现的评价方法,一般以真实性问题情境呈现。表现性任务主要包括设计、赏析、决策、探索四种类型,如课题研究属于探索型的表现性任务。表现性任务综合性程度高,既可以作为学习性评价、学习的评价,也可以作为学习式评价。

#### 4.结构化思维工具

结构化思维工具(也被称为“可视化思维工具”)一般采用图表的方式结构化地呈现思维,主要包括图和表两种形式。这类工具有利于梳理概念与案例、概念与概念的关系,从而让思维结构化,因此,非常适用于大概概念教学。结构化思维工具一般用于学习性评价和学习式评价,但现在也越来越多地用于学习的评价。常见的结构化思维工具包含以下四种。①KWL图表。KWL图表由Ogle于1986年创建,K代表Know,即“(关于主题)我已经知道什么?”(What I know);W代表Want,即“(关于主题)我想知道什么?”(What I want to know);L代表Learned,即“(关于主题)我已经学到了什么?”(What I learned)。<sup>[11]</sup>在单元学习前,可以应用KWL图表了解学生针对相关内容已经知道了什么、想要学习什么。而在单元学习后,则可以应用KWL图表了解学生已经学习了什么,以及未来他们还想学习什么。②思维导图。思维导图(Mind Map)最早由Buzan提出,<sup>[12]</sup>本文将流程图等也归入思维导图中,主要体现主题与主题、概念与概念之间的关系,以从属关系为主。图2为“细胞的生命历程”单元思维导图示例。③概念地图。与思维导图相比,概念地图和思维导图均对结构性的要求较高,但概念地图的难度较大,适用于高年级学生。④个人意涵图(Personal Meaning Mapping,简称“PMM”)。个人意涵图是由福尔克(Falk)开发的一种概念地图的变体,学生可以自由使用单词、短语、句子、图片等描述自己的想法,形式灵活自由,对结构性的要求相对较低。同概念地图一样,个人意涵图的应用需要激活个体经验,反映具体与抽象的关系,能够很好地体现个体差异,但它不像思维导图一般,只呈现概念或只体现

抽象与抽象的关系。图3为个人意涵图示例。<sup>[13]</sup>

表2 单元反思报告



图2 思维导图示例<sup>④</sup>

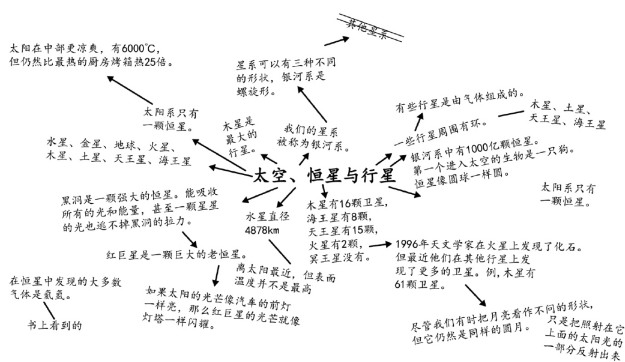


图3 个人意涵图示例

### 5. 自我反思

大概念教学强调“学习式评价”,自我反思也是其中一种重要的评价方法。自我反思以多种形态存在,前文所述的**结构化思维工具可被视为一种正式的自我反思**。例如,让学生对比单元学习前后画的个人意涵图,可以清晰地看到自己的进步,同时也可以和其他同学互学互助。此外,单元反思报告的撰写也是重要的自我反思方式(见表2)。自我反思不仅是一种典型的学习式评价,也可被作为学习性评价,近年来亦被逐渐开发为一种学习的评价,即通过评价学生的反思质量来评定学习水平。特别在项目化学习这类学习中,作品和成果往往由团队合作完成,因此如何评价学生个体的学习水平成了问题,如果仅是让学生自评和互评很难保证学习的评价的“公平性”,因此,以自我反思作为评价方法能够公正地反映每一位学生的水平和增量。

#### (二) 评价连续体的设计

评价是一个连续体,这主要是由**目标的高阶性,以及目标构成的多样化决定的**。如前所述,目标由素养目标、单元大概念、单元具体目标、其他具

反思能力报告

姓名: \_\_\_\_\_

班级: \_\_\_\_\_

学号: \_\_\_\_\_

(团队)反思

1.这节课我(们)学到了什么大概念?

\_\_\_\_\_

2.我(们)觉得这些大概念还可以运用到解决什么问题中?

\_\_\_\_\_

3.对于这些大概念,我(们)还有哪些困惑之处?

\_\_\_\_\_

体单元目标四部分构成。一方面,最重要的素养目标(包含大概念)除了用最终的表现性任务进行评价外,还需要设计一系列过程性的评价。另一方面,要对单元具体目标作梳理和分析,有一些已经体现在表现性评价中,特别是情感维目标、复合技能、高阶思维等,有一些则需要过程中单独进行评价。

评价连续体表现为评价方法的连续体和评价类型的连续体。1.评价方法的连续体。不同的评价目标有不同的评价方法、策略和手段,呈现“连续体”的形态。2.评价类型的连续体。从三种类型的评价来看也是一个连续体,最终的评价一般是“学习的评价”,而过程中学习性评价、学习的评价、学习式评价则穿插于整个学习过程中。一般来说,学习开始时主要以“学习性评价”和“学习式评价”为主,因为需要不断提供反馈。无论是素养、大概念、高阶思维、元认知都需要一个发展的过程,过早地进行“学习的评价”去评定学生不利于他们接受挑战性任务,而“学习的评价”一般在后期出现。但这里要指出的是,“学习的评价”也可能在中期出现,主要是对单一技能和低阶思维进行评价。这两个原因:一是因为这类目标不需要经历一个比较长的学习过程,二是因为单一技能和低阶思维是学习上一级目标的基础,而学生对待“学习的评价”的态度显然是更认真的,在“学习的评价”的督促下对基础知识和技能的掌握会更牢固。图4以“城市”单元为例来展现评价的连续体。<sup>[14]</sup>

从评价的角度来看,单元开始时,引入“城市”相关问题,让学生对日常所处的城市发展概况发生思考,是一种“学习性评价”和“学习式评价”。

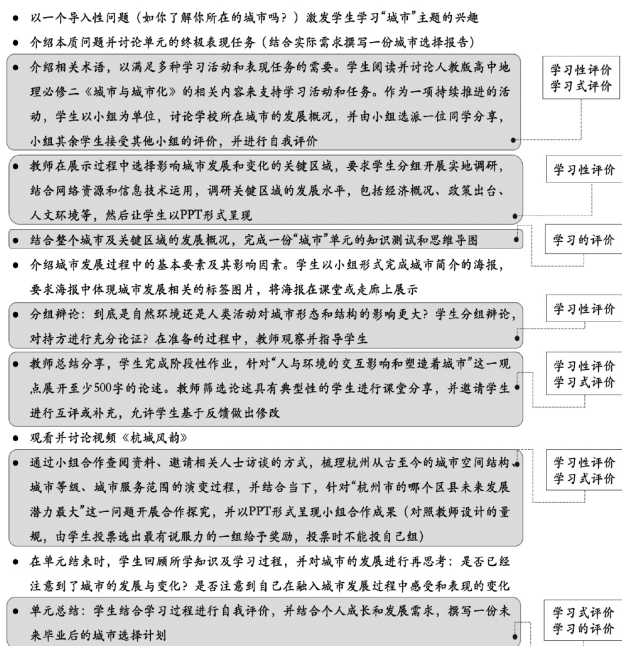


图4 评价连续体示例<sup>⑤</sup>

单元中期，让学生关注城市发展过程中的关键区域，通过查阅资料深入了解城市发展，这是一种贯穿始终的“学习性评价”和“学习式评价”。而“以小组形式完成城市简介的海报→分组辩论，对持方进行充分论证→个人针对‘人与环境的交互影响和塑造着城市’这一观点展开至少500字的论述，并进行互评→小组针对‘杭州市的哪个区县未来发展潜力最大’这一问题开展合作探究，并以PPT形式呈现合作成果”都是过程中有梯度的“学习性评价”，这其中不时安排一些任务需要互评或自评，就是“学习式评价”。单元结束时，每位学生结合个人成长和发展需求，撰写一份未来毕业后的城市选择计划，教师给予评价的同时也请学生进行自评，这既是“学习的评价”，也是“学习式评价”。如前所述，对于素养目标和大概概念一般是到最后再进行“学习的评价”，但中间也会对知识和技能进行“学习的评价”。

注释：

- ①该组样例由浙江大学教育学院徐炜奇设计。
- ②该情境由浙江大学教育学院附属学校孙卓佳设计，杭州滨江教育研究院郑萍指导。
- ③该情境由浙江大学教育学院张倩雯设计。
- ④该思维导图由铜陵市第三中学刘洁设计，铜陵市教育科学研究所何立松指导。
- ⑤该评价连续体由浙江大学教育学院石心玥设计。

参考文献：

- [1]崔允漭.学科核心素养呼唤大单元教学设计[J].上海教育科研,2019,(4).
- [2]易克萨维耶·罗日叶.学校与评估:为了评估学生能力的情境[M].汪凌,周振平,译.上海:华东师范大学出版社,2011:145.
- [3][9]琳达·达令-哈蒙德,弗兰克·亚当森.新一代测评:超越标准化考试,促进21世纪学习[M].韩芳,译.长沙:湖南教育出版社,2018:246,15.
- [4][6]Earl L M. Assessment as learning: using classroom assessment to maximize student learning(2nd ed.)[M]. Thousand Oaks, CA: Corwin, 2013:25~29.
- [5]Stern J, Lauriault N, Ferraro K. Tools for teaching conceptual understanding, elementary: harnessing natural curiosity for learning that transfers[M].Thousand Oaks, CA: Corwin, 2018: 127~128.
- [7]格兰特·威金斯,杰伊·麦克泰格.理解为先模式——单元设计指南(一)[M].盛群力,沈祖芸,柳丰,等译.福州:福建教育出版社,2018:56~57.
- [8][10]格兰特·威金斯,杰伊·麦克泰格.追求理解的教学设计(第二版)[M].闫寒冰,宋雪莲,赖平,译.上海:华东师范大学出版社,2017:18~19,171.
- [11]Ogle D M. KWL: a teaching model that develops active reading of expository text[J]. Reading teacher, 1986, 39(6).
- [12]Buzan T. Mind mapping[M]. Harlow: BBC Active, 2006: 2~18.
- [13]Lelliott A D. Learning about astronomy: a case study exploring how grade 7 and 8 students experience sites of informal learning in South Africa[EB/OL].(2008-02-20)[2021-05-22].https://wiredspace.wits.ac.za/jspui/bitstream/10539/4480/1/LelliottTT.pdf.
- [14]杰伊·麦克泰,格兰特·威金斯.理解为先单元教学设计实例[M].盛群力,等译.宁波:宁波出版社,2019:56.

[徐玲玲 刘徽 曹琦 浙江大学教育学院 310058]